EXHIBIT 2

American Psychological Association

Neuropsychology

In the public domain

2020, Vol. 34, No. 1, 43–52
http://dx.doi.org/10.1037/neu0000569

Comparison of Models of Premorbid IQ Estimation Using the TOPF, OPIE-3, and Barona Equation, With Corrections for the Flynn Effect

Joshua W. Kirton
Evans Army Community Hospital, Fort Carson, Colorado and South Texas Veterans Health Care System, San Antonio, Texas

Jason R. Soble
South Texas Veterans Health Care System, San Antonio, Texas and University of Illinois College of Medicine

Janice C. Marceaux
South Texas Veterans Health Care System, San Antonio, Texas and University of Texas Health Science Center

Johanna Messerly, Kathleen M. Bain, and Troy A. Webber
South Texas Veterans Health Care System, San Antonio, Texas

Chrystal Fullen
South Texas Veterans Health Care System, San Antonio, Texas and Our Lady of the Lake University

W. Alexander Alverson and Karin J. M. McCoy
South Texas Veterans Health Care System, San Antonio, Texas

Objective: Premorbid estimates of intellectual functioning are a key to assessment. This study aimed to compare 3 common measures and assess their accuracy: the Test of Premorbid Functioning (TOPF), Oklahoma Premorbid Intelligence Estimate (OPIE-3), and what is commonly referred to as the *Barona equation*. We also sought to provide appropriate adjustment considering the Flynn effect. ***Method:*** The sample consisted of a cross-section of 189 outpatient veterans receiving neuropsychological assessment including the TOPF and Wechsler Adult Intelligence Scale, 4th ed. (WAIS-IV). Paired sample *t* tests assessed differences between IQ models. Correlations for all models and actual WAIS-IV Full Scale IQ (FSIQ) to establish which model best predicted variance in current IQ. Mean differences were evaluated to establish how closely the models approximated WAIS-IV FSIQ. ***Results:*** The Barona equation estimated higher premorbid IQ than TOPF Simple Demographics Model; however, differences between the models were nonsignificant after a Flynn effect correction for the Barona equation (.23 IQ points per year). The OPIE-3 correlated with FSIQ but overestimated the FSIQ, demonstrating the Flynn effect. TOPF performance models (include word reading) characterized the variance of IQ scores best, but the Flynn-adjusted Barona equation had the smallest mean difference from the actual WAIS-IV FSIQ of any prediction model. ***Conclusion:*** Demographic models for premorbid IQ accurately estimate IQ in adult populations when normed on the test used to measure IQ, or when adjusted for the Flynn effect. A Flynn-corrected Barona score provided a more accurate estimation of WAIS-IV FSIQ than the TOPF or the OPIE-3.

This article was published Online First August 15, 2019.

Joshua W. Kirton, Multidisciplinary-Behavioral Health, Evans Army Community Hospital, Fort Carson, Colorado and Psychology Service (116B), South Texas Veterans Health Care System, San Antonio, Texas; Jason R. Soble, Psychology Service (116B), South Texas Veterans Health Care System and Departments of Psychiatry and Neurology, University of Illinois College of Medicine, Chicago, Illinois; Janice C. Marceaux, Psychology Service (116B), South Texas Veterans Health Care System and Department of Neurology, University of Texas Health Science Center; Johanna Messerly, Kathleen M. Bain, and Troy A. Webber, Psychology Service (116B), South Texas Veterans Health Care System; Chrystal Fullen, Psychology Service (116B), South Texas Veterans Health Care System and Our Lady of the Lake University; W. Alexander Alverson and Karin J. M. McCoy, Psychology Service (116B), South Texas Veterans Health Care System.

The views expressed herein are those of the authors and do not necessarily reflect the views or the official policy of the Department of Veterans Affairs or the U.S. Government. This material is the result of work supported with resources and the use of facilities at the South Texas Veteran's Health Care System.

We would like to acknowledge the veterans who volunteered for this study and offered their data for the betterment of VA care and science.

Correspondence concerning this article should be addressed to Joshua W. Kirton, Multidisciplinary-Behavioral Health, Evans Army Community Hospital, 6541 Specker Road, Building 1830, Fort Carson, CO 80913. E-mail: joshua.w.kirton.civ@mail.mil

General Scientific Summary
Well known in the measurement of IQ, the Flynn effect presents an obstacle for use of the same instruments over time. This article examined the presence of the Flynn effect in IQ estimation models and evaluates their predictive value for WAIS-IV scores. We found that the Barona equation could be corrected with a constant inflation rate to accurately predict WAIS-IV scores with no significant difference from that of the TOPF.

Keywords: premorbid IQ, premorbid intelligence, estimation of IQ, modeling premorbid IQ, Flynn effect

Supplemental materials: http://dx.doi.org/10.1037/neu0000569.supp

A critical component of psychological and neuropsychological evaluation involves comparison of test performance with estimates of overall intellectual ability (e.g., IQ) to determine whether assessment results reflect a change from premorbid IQ. Efforts have long been directed at the development and validation of brief measures, which facilitate a practical approach to estimating premorbid IQ, including demographics-based models (Barona, Reynolds, & Chastain, 1984), "hold" tests (Nelson & Willison, 1991; Wechsler, 2001), "best-performance" models (Vanderploeg, Schinka, & Axelrod, 1996), and a combination of demographics and "hold" test data (Schoenberg, Scott, Duff, & Adams, 2002; Pearson, 2009).

Initially, demographics-based models of estimated premorbid IQ relied on education (Fogel, 1964; Ladd, 1964), age, race, and sex (Reynolds & Gutkin, 1979) as placeholders for IQ, with education and race most strongly predictive of Full Scale IQ (FSIQ) on the original Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955; Wilson et al., 1978). These proxy indices were hypothesized to reflect quality and access to intellectual resources throughout development, although several related variables (e.g., geographic region of residence) were not investigated. Subsequently, Barona et al. (1984) showed that age, sex, race, education, occupational status, geographic region of residence, and urban–rural residence predicted 36% of the variance in FSIQ on the revised WAIS (WAIS–R; Wechsler, 1981) and produced the following equation, typically called the *Barona equation*, to estimate premorbid intellectual functioning: 54.96 + 0.47(age) + 1.76(sex) + 4.71(race) + 5.02(education) + 1.89(occupation) + 0.59(region).[1] Although an attractive, cost-effective method given the equation is freely available and uses information collected during a clinical interview, a primary limitation to this equation, and other demographic models, is susceptibility to the Flynn effect (Flynn, 1987)—the tendency of average scores on an intelligence test to increase over time. As such, any given demographics model developed with IQ estimates from a previously normed test would likely increasingly overestimate premorbid IQ over time (Norton, Watt, Gow, & Crowe, 2016). Application of the Barona equation for prediction of WAIS, 4th ed. (WAIS-IV; Wechsler, 2008) FSIQ (Batchelor & Meyers, 2013) showed only modest correlation (r = .31) with FSIQ scores.

Alternative approaches, to the Barona equation, rely solely on testing performance to approximate premorbid IQ. Putative "hold" tests—measures of constructs thought to be relatively unaffected by neurocognitive decline—typically assess word reading ability and are based on the notion that this ability is strongly related to overall intellectual abilities and allows access to previous knowledge and education, without taxing other cognitive abilities (Willshire, Kinsella, & Prior, 1991). David Wechsler first developed performance-based estimation using the WAIS Vocabulary subtest, with the idea that it would be robust to degeneration, and developed formulas to compare the Vocabulary subtest to other verbal scores to obtain a degeneration index score. This method was further examined and progressed by other researchers who established "best-performance" models using Vocabulary and Picture Completion subtests from previous WAIS–R and WAIS, 3rd ed. (WAIS-III; Wechsler, 1997; Krull, Scott, & Sherer, 1995; Lezak, Howieson, & Loring, 2004; McFie, 1975; Vanderploeg & Schinka, 1995; Wechsler, 1981, 1997). Several word reading tests, such as the Wide Range Achievement Test (Wilkinson & Robertson, 2006) and National Adult Reading Test (Nelson & Willison, 1991), were developed and became increasingly popular among neuropsychologists (Smith-Seemiller, Franzen, Burgess, & Prieto, 1997) given evidence that they are similarly resistant to multiple neurological conditions (Carlozzi et al., 2011; Maddrey, Cullum, Weiner, & Filley, 1996) and are minimally burdensome to patients. However, several studies showed that word reading ability may decline with progression of some neurodegenerative diseases, such as Alzheimer's disease (Ashendorf, Jefferson, Green, & Stern, 2009; Cockburn, Keene, Hope, & Smith, 2000; Taylor, 2000) and Huntington's disease (O'Rourke et al., 2011). Thus, the robustness of performance models is tied to the appropriateness of the underlying assumptions and the appropriateness of the data upon which it relies to derive the estimate (Lezak et al., 2004).

The "best-performance" approach is based on the highest test scores, performance in everyday tasks, observable behavior, or as evidenced by life achievements (Lezak et al., 2004). Although some evidence suggests that this method accurately estimates premorbid IQ across clinical groups (Vanderploeg et al., 1996), other data suggest this method overestimates premorbid IQ (Mortensen, Gade, & Reinisch, 1991), particularly when based on highest test scores that may not represent the full range of abilities.

Rather than relying entirely on demographic predictions or "hold" tests, several researchers have developed measures of pre-

[1] This model uses the following coding of demographic variables: age (16–17 = 1, 18–19 = 2, 20–24 = 3, 25–34 = 4, 35–44 = 5, 45–54 = 6, 55–64 = 7, 65–69 = 8, 70+ = 9), sex (female = 1, male = 2), race (Black = 1, other = 2, White = 3), education (0–7 years = 1, 8 = 2, 9–11 = 3, 12 = 4, 13–15 = 5, ≥16 = 6), occupational status (1 = unskilled worker, 2 = farm laborer, 3 = semiskilled laborer, 4 = not in labor force, 5 = skilled worker, 6 = managerial/sales), geographic region of residence (South = 1, North-Central = 2, West = 3, Northeast = 4), urban–rural residence (rural = 1, urban = 2).

morbid IQ that combine both approaches. For example, the Oklahoma Premorbid Intelligence Estimate (OPIE-3; Krull et al., 1995) was developed to predict WAIS–R FSIQ, Verbal IQ (VIQ), and Performance IQ (PIQ) using both demographic data (i.e., age, education, occupation, and gender) and current performance on a subset of WAIS–R subtests. The Oklahoma Premorbid Intelligence Estimate-3 (OPIE-3) was later developed for prediction of WAIS-III FSIQ rather than the WAIS–R (Schoenberg et al., 2002) and demonstrated excellent concurrent validity in a normal population (i.e., the WAIS-III standardization sample). The OPIE-3 consists of several formulas that can be calculated, and a best performance procedure can be followed looking at the highest result among the formulas. The measure was not without limitations; most notably, the use of subtests from the WAIS-III in estimating FSIQ introduced psychometric concerns including regression to the mean at the extremes of the distribution, and the four-subtest formula is thought to be subject to variability among impaired samples because of its reliance on tests that can be affected by degeneration (Schoenberg et al., 2002). The OPIE(2ST) formula did not systematically over- or underestimate IQ in a cognitively intact sample but was more likely than the four-subtest formula to be more than five points from actual IQ (Schoenberg, Duff, Scott, Patton, & Adams, 2006). Overall, the OPIE-3 generally performed well in a normal sample, particularly in the middle of the distribution but tends to underestimate IQ for young and old, and overestimate IQ in the middle age ranges (Schoenberg et al., 2006). In another study the OPIE-3 overestimated IQ in 78–92% of patients with the OPIE-3 (Best) formula being the worst offender (61% were overestimated by 10 IQ points), and the two and four subtest formulas being the most stable and closest to actual IQ (Schoenberg, Duff, Scott, & Adams, 2003).

More recently, Pearson (2009) introduced a measure of estimated WAIS-IV premorbid IQ, the Test of Premorbid Functioning (TOPF), which uses a "hold" test of irregular word reading ability in combination with demographic information. Clinicians can produce estimated FSIQ scores based on demographics alone, word reading performance alone, or a combination of the two. Notably, scores on the TOPF are calculated with proprietary formulae, and the regression equations for demographic-based FSIQ predictions are not published. Therefore, use of the TOPF entails an inherent cost not only for test materials, but also for access to scoring software.

Past comparisons of the various approaches to IQ estimation have indicated that all have similar validity for predicting WAIS FSIQ (Axelrod, Vanderploeg, & Schinka, 1999 [WAIS–R]; Spinks et al., 2009 [WAIS-III]), with the caveat that approaches which involve more testing generally produce estimates which are more accurate in normal samples, but may not perform well in impaired samples (Larrabee, Largen, & Levin, 1985; Ashendorf et al., 2009; Cockburn et al., 2000; Taylor, 2000; O'Rourke et al., 2011). Another consistent finding has been poor prediction of premorbid IQ for individuals within the tails of the normative distribution (Franzen, Burgess, & Smith-Seemiller, 1997; Spinks et al., 2009). Given similar concurrent validity, clinicians can ultimately select a preferred premorbid IQ estimation approach based on the characteristics of the patient and/or the constraints of the situation. Because time often is at a premium in neuropsychological evaluations, premorbid IQ estimation approaches—which are the least time-consuming may be preferred. Cost of test materials and, by extension, overall evaluation costs and health care expenditure is another important consideration. Finally, patients have differing levels of tolerance for formal testing, and therefore may be more or less appropriate for approaches that require test administration as opposed to those that rely primarily on clinical interview.

One study examined demographic and reading based performance models to estimate premorbid IQ and compared them to each other (Norton et al., 2016). These analyses demonstrated that formulas based on previous forms of the WAIS demonstrated the Flynn effect of inflated IQ estimations over time. However, the major weakness of this study was that it only compared methods of IQ estimation and did not compare the results to current IQ with a complete WAIS score. The current study had several aims. The first was to test the accuracy of selected measures of premorbid intellectual functioning that use demographics models only (i.e., Barona equation andTOPF) for predicting WAIS-IV FSIQ in a clinical sample. The second was to compare the accuracy of the Barona, OPIE 3, and TOPF approaches to determine whether any one approach (i.e., Barona, OPIE 3, and TOPF) performed significantly better than the others in predicting actual IQ in unimpaired individuals. Third was to delineate the accuracy of the different TOPF models (i.e., TOPF demographics alone, TOPF word reading performance alone, or a combination of the two) in estimating IQ in unimpaired individuals. Finally, the current study sought to develop and validate a Flynn corrected Barona equation for predicting WAIS-IV FSIQ that accounts for the Flynn effect and validate it against actual FSIQ and the TOPF in unimpaired individuals. Ultimately, the thrust of the current study was to address a key limitation in the existing literature and provide additional information which can guide decision making as clinicians approach estimation of premorbid IQ for adult patients in the applied setting.

Method

Participants

This cross-sectional study used data collected as part of an ongoing, institutional review board-approved clinical research project with veterans who received neuropsychological evaluation at a VA Medical Center from 2011 to 2016. All participants provided written informed consent to participate in the study following completion of their clinical evaluation. Sample consisted of individuals presenting for outpatient neuropsychological evaluation. Participants who completed the TOPF (Pearson, 2009) and WAIS-IV (Wechsler, 2008) were included. One participant was excluded from analyses as he did not fit into one of the three racial categories analyzed, and anyone failing ≥2 performance validity tests (PVTs) was excluded (N = 10). PVTs varied between patients and providers but always included at least three different PVTs. We used failures on ≥2 PVTs across multiple PVTs administered throughout the testing battery as the exclusion criteria for invalid performance because burgeoning evidence suggests ≥2 PVT failures is an acceptable standard of practice in the field for determining invalidity (Boone, 2009). This reflects the well-documented notion that 1 PVT failure represents a single data point and may reflect the impact of multitude of factors (Erdodi & Lichtenstein, 2017; Larrabee, 2003). Particularly considering evidence suggesting that 1 PVT failure is common among valid-

cognitively impaired examinees in mixed clinical samples (such as the current sample) with increased likelihood of cognitive impairment (Schroeder, Martin, Heinrichs, & Baade, 2018; Webber, Critchfield, & Soble, in press), we used a cutoff of ≥2 PVT failures (rather than ≥1 PVT failures) to exclude participants. The final sample consisted of 189 veterans (86% male; $n = 164$) with an average age of 60.75 years ($SD = 12.27$; range = 20–84) and average education of 13.46 years ($SD = 2.97$; range = 6–20). Racial composition was diverse such that 51.85% were White ($n = 98$), 37.6% Hispanic ($n = 71$), and 10.6% African American ($n = 20$). English/Spanish bilingual patients made up 30.2% of the sample ($n = 57$), and remaining participants were monolingual English speakers. All patients were fluent in English and tested in English.

Board-certified clinical neuropsychologists selected neuropsychological test batteries and norms for each participant based on demographic considerations and referral questions. Batteries were not identical across participants, but all included measures to assess general intellectual functioning, attention, processing speed, visuospatial skills, language, memory, and executive functioning. Cognitively unimpaired veterans made up 42.8% ($n = 81$) of the sample, whereas 57.1% ($n = 108$) met formal *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (*DSM–5*; American Psychiatric Association, 2013) diagnostic criteria for a neurocognitive disorder. Among the cognitively unimpaired, $n = 16$ (8.5%) had no diagnosis, and $n = 65$ (34.4%) had psychiatric diagnoses of depression ($n = 28$), anxiety/PTSD ($n = 30$), or other ($n = 7$; substance use, somatization, bipolar, psychosis, other). For those with a neurocognitive disorder, the most common etiology was stroke/vascular disease ($n = 36$; 19%). Additional etiologies included Alzheimer's disease ($n = 13$; 6.9%), multiple etiologies ($n = 29$; 15.3%), epilepsy ($n = 2$; 1.1%), traumatic brain injury ($n = 1$; .5%), frontotemporal dementia ($n = 3$; 1.6%), Parkinson's disease ($n = 5$; 2.6%), Parkinson's plus ($n = 1$; .5%), and medication-induced ($n = 1$; .5%). Impaired individuals were only used in analyses comparing demographics models to avoid the confounding effect of cognitive impairment on performance-based models. Some unimpaired individuals also did not have complete WAIS-IV scores, and thus were not included in analyses for the WAIS-IV and only in the analyses for modeling demographics (see Tables 3 and 4).

Measures

OPIE-3. The OPIE-3 (Schoenberg et al., 2002) provides several calculations combining WAIS-III subtest raw scores and demographic variables to estimate premorbid WAIS-III FSIQ scores. Five separate algorithms were developed by examining data from the WAIS-III standardization sample. Each algorithm uses a unique combination of WAIS-III subtests (i.e., Vocabulary, Information, Matrix Reasoning, and Picture Completion) and demographic data including age, education, gender, ethnicity, and region to predict FSIQ. The four-subtest equation [OPIE-3(4ST)] included all the aforementioned WAIS-III subtests in predicting FSIQ. The two-subtest equation [OPIE-3(2ST)] included only Vocabulary and Matrix Reasoning. A Verbal subtest equation was generated, but it was found that the Information subtest did not contribute significant variance over and above the Vocabulary subtest, thus Vocabulary is the only subtest in the OPIE-3V equation. The OPIE-3P equation used both nonverbal subtests (i.e., Matrix Reasoning, Picture Completion) in the prediction of FSIQ. A final prediction equation used only Matrix Reasoning and demographic variables (OPIE-3MR). Each OPIE-3 equation provided an estimate of FSIQ that did not differ significantly from the actual FSIQ of the validation sample, and all were significantly correlated with actual FSIQ, with OPIE-3(4ST) and OPIE-3(2ST) resulting in the highest correlations ($r = .924$ and .897, respectively).

The algorithms were clinically validated (Schoenberg et al., 2002) and found to significantly overestimate actual FSIQ in patients with brain injuries, as expected. The OPIE-3(4ST) was noted to be a better estimate of current, rather than premorbid FSIQ, in patient groups. In groups of patients with diffuse or right lateralizing lesions, OPIE-3V is the recommended method for predicting premorbid FSIQ. Conversely, for patients with left lateralizing lesions, OPIE-3MR is recommended as the best method for predicting premorbid FSIQ. The authors presented data for the OPIE-3(Best), a guideline for selection of OPIE-3V, OPIE-3MR, or OPIE-3(2ST) as the best premorbid estimate of FSIQ depending on the pattern of subtest performance. Specifically, in the case of a discrepancy of at least one point between Vocabulary and Matrix Reasoning age-corrected scaled scores, the algorithm using the higher of the two subtests is thought to reflect the best estimate of premorbid FSIQ; in the case of equivalent age-corrected scaled scores for Vocabulary and Matrix Reasoning, OPIE-3(2ST) is recommended. In the current study, we used WAIS-IV test scores to calculate the OPIE-3(2ST), OPIE-3(V), and OPIES-3(MR). Although it is expected that the use of WAIS-IV scores will attenuate any findings of the Flynn effect, it is anticipated that using the formula betas will still establish the presence, but not the extent of the Flynn effect, in overestimating premorbid IQ using formulas normed to previous versions of the WAIS.

Barona. Barona and colleagues (1984) used data from the WAIS–R standardization sample to predict VIQ, PIQ, and Full-Scale IQ (FSIQ) from demographic variables. Demographic predictors included age, sex, race, geographic region, occupation, education, and urban–rural residence. All of these variables contributed significantly to total explained variance across the three IQs, though urban–rural residence was not significantly predictive of VIQ, and sex was not significantly predictive of PIQ. Education, race, and occupation were found to be the most significant predictors overall. The authors commented on the benefit of premorbid FSIQ estimation that does not rely on test performance, which is likely to be compromised in patients with brain injuries, although they did not provide validation of their equations. As discussed by these authors, a limitation of the regression equation is the likelihood of artificially lowering or raising estimated scores for individual cases falling at the extremes of actual functioning, due to regression to the mean. We altered coding from the original Barona equation slightly for this study including for ethnicity/race: African American (1), Hispanic (2), and Caucasian (3); as well as age 70+ rather than 70–74 ($n = 52$; range = 70–84; $M = 74.63$; Median = 73; $SD = 4.22$).

WAIS-IV. The WAIS-IV is a well validated and gold standard measure of concurrent intellectual functioning (Wechsler, 2008). It is often considered the gold standard in measurement of IQ.

Advanced Clinical Solutions TOPF. The TOPF (Pearson, 2009) is designed to provide an estimate of premorbid IQ and was created as a revision of the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). The test uses a single word reading paradigm including 70 irregular words and relies on the relative resilience of word recognition and decoding to the effects of brain injury and neurological compromise. Revisions to WTAR were also designed to improve the predictive range of scores for the TOPF. These included expanding the upper range of education level, adding occupation to the demographic model, and transformation of the TOPF age-adjusted standard score into WAIS-IV and Wechsler Memory Scale, 4th ed. equated scores, to reduce the impact of regression to the mean. The TOPF provides three general predictive models: (a) the simple demographic model, using sex, ethnicity, education level, region of residence, occupational level; (b) the complex demographic model, accounting for additional personal and developmental variables including amount of sleep the night before testing, perceived wealth in the neighborhood of residence, social activity, exercise level, elementary school quality, perceived wealth of childhood neighborhood, parents' levels of education, and parents' occupational levels; and (c) the TOPF-equated model, applying demographic variables in addition to word reading performance to estimate premorbid IQ. Based on standardization of the TOPF, the simple demographics model tends to adequately predict premorbid IQ with relatively little information. Additional variables can be considered in the complex demographics model, particularly if the clinician has reason to believe that simple demographic factors may not capture relevant information about the individual. The TOPF-equated model provides additional predictive precision when some demographic information is unavailable or does not capture personal factors well.

Data Analyses

Paired samples *t* tests were used to test for differences between the TOPF simple demographics model and the standard Barona equation for the overall sample, and then separately by racial/ethnic groups (i.e., White, Hispanic, African American) and age bands (i.e., ≤50, 51–65, 66+), to establish that there are no significant differences in the performance of the equations that use only demographics across groups. Analyses were then repeated to evaluate two corrections for the Flynn effect (Norton et al., 2016) in the Barona equation given that this equation, but not the TOPF, relies on previous versions of the WAIS and requires correction for the Flynn effect (See Supplement 1 in the online supplemental material); specifically, we used a 0.3 and 0.23 IQ points (Trahan, Stuebing, Fletcher, & Hiscock, 2014) per year inflation correction for the period between the publication of the norms for the version of the WAIS that each respective model was developed to predict. We used the publication dates as proxies for the years the tests were normed. As we are predicting WAIS-IV scores, a correction need only be applied that adjusts to the normative sampling for the WAIS-IV resulting in a constant to predict WAIS-IV scores using the Barona. These correction rates over the 27 years between the publication of WAIS–R (Wechsler, 1981) and WAIS-IV (Wechsler, 2008) resulted in −8.1 and −6.21 IQ point corrections for the Flynn effect at a high and low point, respectively. Thus, to calculate the Flynn effect-corrected Barona Demographics Model, we used the original equation (Barona et al., 1984), minus the corrections to obtain two different corrected estimated WAIS-IV scores from the Barona equation. These corrected scores were compared to the TOPF Simple Demographics Model. Next, we excluded cognitively impaired participants and examined correlations for the WAIS-IV FSIQ and all demographic- and performance-based IQ estimates. This allowed us to evaluate pairwise variance comparisons and establish the premorbid IQ models most highly correlated with current IQ. Four paired samples *t* tests were completed, three to compare the corrected Barona Demographics Model to WAIS-IV FSIQ, Best TOPF model (Dem + TOPF), and the OPIE-3(2ST) which was the highest correlate with IQ; and a fourth comparing the means of the OPIE-3(2st) and WAIS-IV FSIQ. This was to establish if, despite correlation, there were significant differences in group means. Finally, to control for false positive rate associated with multiple *t* tests, the false discovery rate (FDR) procedure with a .05 maximum false discovery rate was applied to all *t* test comparisons (Benjamini & Hochberg, 1995; Glickman, Rao, & Schultz, 2014). FDR corrections were not used for correlations given evidence for reduced accuracy and overcorrection of highly correlated data (Farcomeni, 2006; Zhang & Coombes, 2012), as is the case in the current study.

Results

Regarding the demographics models, for the final total sample (N = 189), the TOPF Simple Demographics Model and Barona equation were significantly different with the Barona equation estimating higher premorbid IQs than the TOPF Simple Demographics model (see Table 1). After correcting for the Flynn effect on the Barona equation, the .3 point per year corrected Barona equation continued to show a significant difference from the TOPF Simple Demographics model with the .3 point per year correction on the Barona overcorrecting, while nonsignificant differences were found between the .23 point per year Barona and TOPF Simple Demographics Model. Among racial groups, a similar pattern emerged in which significant differences were initially present when comparing the original Barona model to the TOPF Simple Demographics Model across ethnicity/races, but nonsignificant differences were found between models after the .23 point per year correction was implemented for the Barona demonstrating no significant difference in performance between groups and the presence of Flynn effect in predicting individual racial/ethnic group scores. The same pattern was observed across age groups where prior to correction, the Barona overestimated higher IQ, but with correction for the Flynn effect (.23 IQ points per year), there were nonsignificant differences between TOPF Simple Demographics Model and the corrected Barona across age groups. Thus, the Flynn effect is present in demographic IQ estimation when using older equations, but this can be easily corrected by using the calculated Flynn effect shown in the meta-analysis by Trahan et al., (2014). The lack of significant findings by race/ethnic group demonstrates that the Flynn adjusted Barona equation (.23/year) performs no differently for any one group than the TOPF Simple Demographics Model.

Correlations between actual WAIS-IV scores for cognitively normal participants and various predictive models are available in Table 2. This shows that OPIE-3 (2ST) model best correlated with WAIS-IV FSIQ, and that in general performance-based measures were more highly correlated to WAIS-IV FSIQ. This can also be seen when looking at the three TOPF models. The Demograph-

Table 1
t-Tests Comparing the Barona Equation to the TOPF Demographics (N = 189)

Demographic	*n*	TOPF *M* (*SD*)	Barona *M* (*SD*)	*df*	*t*	Cohen's *d*
Total sample	189	98.90 (10.31)	104.89 (9.15)	188	−14.91***	.615
Caucasian	98	104.40 (8.20)	109.57 (7.31)	97	−9.53***	.666
Hispanic	71	93.76 (8.04)	100.21 (7.92)	70	−10.3***	.808
African American	20	90.20 (11.69)	98.57 (9.37)	19	−5.77***	.790
Age						
≤50	31	96.23 (9.55)	101.37 (7.23)	30	−5.039***	.607
51–65	82	98.39 (10.03)	104.42 (8.46)	81	−10.19***	.650
66+	76	100.54 (10.75)	106.83 (10.12)	75	−9.67***	.603
With .3 IQ point/year correction (8.1) to Barona						
Total sample	189	98.90 (10.31)	96.79 (9.15)	188	5.25***	.216
Caucasian	98	104.40 (8.20)	101.47 (7.31)	97	5.4***	.377
Hispanic	71	93.76 (8.04)	92.11 (7.92)	70	2.629	.207
African American	20	90.20 (11.69)	90.47 (9.37)	19	−.184	.025
Age						
≤50	31	96.23 (9.55)	93.27 (7.23)	30	2.89	.349
51–65	82	98.39 (10.03)	96.32 (10.03)	81	3.5*	.223
66+	76	100.54 (10.75)	98.73 (10.12)	75	2.78	.173
With .23 IQ point/year correction (6.21) to Barona						
Total sample	189	98.90 (10.31)	98.68 (9.15)	188	.547	.023
Caucasian	98	104.4 (8.20)	103.36 (7.31)	97	1.918	.134
Hispanic	71	93.76 (8.04)	94.00 (7.92)	70	−.39	.030
African American	20	90.20 (11.69)	92.36 (9.37)	19	−1.49	.203
Age	31	96.23 (9.55)	95.16 (7.23)	30	1.043	.126
≤50						
51–65	82	98.39 (10.03)	98.21 (8.46)	81	.304	.019
66+	76	100.54 (10.75)	100.62 (10.12)	75	1.38	.008

Note. TOPF = Test of Premorbid Functioning. All *p* values reflect false discovery rate (FDR)-corrected *p* values.
* $p < .01$. *** $p < .0001$.

ics + TOPF model was the most highly correlated with WAIS-IV FSIQ scores, followed by TOPF-Equated model, and then by Simple Demographics Model.

Despite high correlations with the WAIS-IV FSIQ, further examination of mean differences in the cognitively unimpaired subsample (see Table 4) found the adjusted Barona equations had the smallest mean difference. However, when examining the means (see Table 3) for significant differences (see Table 4) in the cognitively unimpaired subsample, it was shown that, despite high correlation with WAIS-IV FSIQ, the group means were significantly different. Individually, R^2 was calculated in Table 2 and the variance of FSIQ was best approximated by the OPIE-3 (2ST), but that the OPIE-3 overestimated IQ as a group due to being created for a previous version of the WAIS because of the Flynn effect. Furthermore, the corrected Barona did not significantly differ from measured WAIS-IV FSIQ in unimpaired individuals.

Although there was a significant difference between the performance models and the corrected Barona model, there were nonsignificant differences between the corrected Barona model and TOPF Demographics model and actual WAIS-IV FSIQ. All the OPIE-3 models overpredicted WAIS-IV FSIQ, further supporting the idea that models created for earlier versions of the WAIS demonstrate the Flynn effect and need to either be corrected or not used to estimate WAIS-IV scores.

Discussion

The current study sought to address a key gap in the current literature on premorbid IQ estimation and respond to limitations of existing literature examining the clinical utility of several common premorbid IQ estimate measures (i.e., TOPF, Barona equation, and OPIE-3). These measures encompass the current scope of what is used to estimate premorbid IQ including demographic models, "hold" tests (e.g., word reading), combination "hold"/demographic, and "best-performance" models. Specifically, we aimed to validate these measures with a WAIS-IV FSIQ in a diverse mixed clinical sample, to determine which model, is more accurate in predicting IQ, to investigate the presence of the Flynn effect in the models, and to produce a corrected Barona equation to account for the Flynn effect in this sample.

Three approaches (i.e., OPIE-3, TOPF, and Barona) for predicting WAIS FSIQ among cognitively intact individuals were examined with respect to their relationship to current WAIS-IV FSIQ. The OPIE-3 models demonstrated the strongest correlations with WAIS-IV FSIQ, which was not unexpected given the OPIE-3 uses subsets of the WAIS-IV to predict the total WAIS-IV score, thus sharing a portion of the variance (Krull et al., 1995; Schoenberg et al., 2002), and likely highlighting issues of multicollinearity and regression to the mean. Despite the original paper (Schoenberg et al., 2002) suggesting that multicollinearity was not significant, we

Table 2
Correlations and R^2 Among IQ and IQ Estimates for the No Cognitive Impairment Group

Variable	1	2	3	4	5	6	7	8	9	10	11
r											
1. WAIS-IV FSIQ	1	.936**	.543**	.720**	.729**	.872**	.832**	.765**	.567**	.567**	.567**
2. WAIS-IV GAI	.936**	1	.497**	.703**	.700**	.888**	.818**	.778**	.491**	.491**	.491**
3. TOPF Simple Dem.	.543**	.497**	1	.420**	.812**	.509**	.519**	.608**	.883**	.883**	.883**
4. TOPF Equated	.720**	.703**	.420**	1	.831**	.722**	.636**	.577**	.448**	.448**	.448**
5. Dem. + TOPF	.729**	.700**	.812**	.831**	1	.707**	.681**	.671**	.779**	.779**	.779**
6. OPIE3_2ST	.872**	.888**	.509**	.722**	.707**	1	.898**	.863**	.513**	.513**	.513**
7. OPIE3_V	.832**	.818**	.519**	.636**	.681**	.898**	1	.649**	.578**	.578**	.578**
8. OPIE3_MR	.765**	.778**	.608**	.577**	.671**	.863**	.649**	1	.643**	.643**	.643**
9. Barona FSIQ	.567**	.491**	.883**	.448**	.779**	.513**	.578**	.643**	1	1.000**	1.000**
10. Barona Flyn High	.567**	.491**	.883**	.448**	.779**	.513**	.578**	.643**	1.000**	1	1.000**
11. Barona Flyn Low	.567**	.491**	.883**	.448**	.779**	.513**	.578**	.643**	1.000**	1.000**	1
R^2											
1. WAIS-IV FSIQ	1	.876	.294	.518	.531	.760	.692	.585	.321	.321	.321
2. WAIS-IV GAI	.876	1	.247	.494	.49	.788	.669	.605	.241	.241	.241
3. TOPF Simple Dem.	.294	.247	1	.176	.659	.259	.269	.369	.779	.779	.779
4. TOPF Equated	.518	.494	.176	1	.690	.521	.404	.332	.200	.200	.200
5. Dem. + TOPF	.531	.49	.659	.690	1	.499	.463	.450	.606	.606	.606
6. OPIE3_2ST	.760	.788	.259	.521	.499	1	.806	.744	.263	.263	.263
7. OPIE3_V	.692	.669	.269	.404	.463	.806	1	.421	.334	.334	.334
8. OPIE3_MR	.585	.605	.369	.332	.450	.744	.421	1	.413	.413	.413
9. Barona FSIQ	.321	.241	.779	.200	.606	.263	.334	.413	1	1	1
10. Barona Flyn High	.321	.241	.779	.200	.606	.263	.334	.413	1	1	1
11. Barona Flyn Low	.321	.241	.779	.200	.606	.263	.334	.413	1	1	1

Note. 2ST = two subtests; Dem. = Demographics; FSIQ = Full Scale Intelligence Quotient; MR = Matrix Reasoning; OPIE = Oklahoma Premorbid IQ Estimate; TOPF = Test of Premorbid Functioning; V = Vocabulary; WAIS-IV = Wechsler Adult Intelligence Scale, 4th ed.
** Correlation is significant at the .01 level (2-tailed).

assessed the correlation of predictor variables and found there to be significantly high enough correlations to make multicollinearity a concern for the OPIE-3(2ST). Although the OPIE-3 models that only use one subtest do not have problematic multicollinearity, the shared variance and presence of the Flynn effect invalidate these models for use without adjustment. Regarding the three TOPF models, a pattern was observed in which the combination of a performance-based measure and demographic estimates had the strongest correlation with the FSIQ, followed by the TOPF-Equated (word reading) model. Finally, models that only included demographic information (i.e., TOPF Simple Demographics Model and the Barona equation) were moderately correlated with WAIS-IV FSIQ. Although the performance-based models were more highly correlated with actual FSIQ, these findings may reflect issues of multicollinearity and lack of independence in the OPIE-3 and other performance-based models. Although these performance models have similar variances this is likely due to the multicollinearity. The shared variance is of less concern given that the performance model of the OPIE is systematically overpredicting IQ. Thus, the ideal measure is one that not only has correlating variance, but a low mean difference from actual IQ.

Table 3
Mean Differences for Premorbid IQ Estimates Compared to Mean WAIS-IV FSIQ for No Cognitive Impairment Group (N = 73)

Measure	*M*	*SD*	Mean difference from WAIS-IV	*n*
WAIS-IV Full Scale IQ	98.68	10.64	—	65
WAIS-IV General Abilities Index	100.88	10.64	2.2	66
TOPF Simple Dem.	99.96	10.04	1.28	73
TOPF Equated	99.64	10.41	.96	72
Dem. TOPF	100.5	8.79	1.82	72
OPIE3_2ST	104.39	10.29	5.71	55
OPIE3_V	100.68	9.74	2	56
OPIE3_MR	108.24	9.4424	9.56	62
Barona FSIQ	104.74	9.07	6.06	73
Barona_Flynn_High (.3/yr.)	96.64	9.07	−2.04	73
Barona_Flynn_Low (.23/yr.)	98.53	9.07	−.15	73

Note. 2ST = two subtests; Dem. = Demographics; FSIQ = Full Scale Intelligence Quotient; MR = Matrix Reasoning; OPIE = Oklahoma Premorbid IQ Estimate; TOPF = Test of Premorbid Functioning; V = Vocabulary; WAIS-IV = Wechsler Adult Intelligence Scale, 4th ed.

Regarding the latter objective, all three approaches (i.e., OPIE-3, TOPF, Barona) were examined with respect to their ability to predict current WAIS-IV FSIQ. The Flynn effect was observed, as the original Barona equation and the OPIE-3 (2ST) consistently overpredicted FSIQ. These findings not only continue to demonstrate the need for clinicians to be mindful of the Flynn effect when selecting measures in general, but also specifically demonstrated that these original equations are not recommended for current premorbid estimation. The clinical implication of continuing to use the unadjusted original Barona equation or the OPIE 3 (2ST) is that an individual's performance on cognitive measures may appear to reflect a decline from premorbid functioning and be misattributed as an impairment rather than an inflated premorbid IQ estimate (Norton et al., 2016). Consequently, accounting for the Flynn effect should be incorporated into a clinician's premorbid IQ estimate choice.

Table 4
Paired Sample T-Tests for Select Group Mean Differences Among the No Cognitive Impairment Group (N = 72)

Group 1	Group 2	*n*	*df*	*t*	Cohen's *d*
Corrected Barona 98.15 (8.88)	WAIS-IV FSIQ 99.10 (10.89)	58	57	−.78	−.096
Corrected Barona 98.56 (9.13)	Best TOPF model (Dem + TOPF) 100.5 (8.80)	72	71	−2.76*	−.216
Corrected Barona 98.42 (9.11)	Best Predictive Model (OPIE 2ST) 104.39 (10.29)	55	54	−4.60***	−.614

Note. 2ST = two subtests; Dem. = Demographics; FSIQ = Full Scale Intelligence Quotient; OPIE = Oklahoma Premorbid IQ Estimate; TOPF = Test of Premorbid Functioning; WAIS-IV = Wechsler Adult Intelligence Scale, 4th ed.
* $p < .01$. *** $p < .0001$.

Notably, this study demonstrated that the Flynn effect was controlled when a correction of .23 IQ points per year was subtracted from the original Barona equation, further validating the meta-analysis by which this correction was derived (Trahan et al., 2014). In the Trahan article, the authors show the .23 inflation over time, but they hedge their findings performing additional analyses and acknowledging that this is lower than what Flynn originally reported. Their final conclusion is that the .23 correction is best used when correcting for earlier editions of the WAIS, and .3 for later editions of the WAIS, thus our study validates the use of a .23-point correction for models originally predicting WAIS–R scores such as the Barona. Furthermore, if one's regular practice is to obtain an actual current IQ (adjusted for Flynn related inflation), rather than a WAIS-IV prediction, a .23 correction should be subtracted from both the obtained WAIS-IV IQ score and Barona estimate. This becomes necessary as it would be anticipated that inflation has occurred in the current edition of the WAIS as well and emphasizes the importance of regular standardization or revamping of IQ tests about once a decade. The inflation of .23–.3 points sums to about 2.3–3 points per decade, which is at the edge of the standard error of measurement for the WAIS.

In a direct comparison between the corrected Barona equation and the TOPF Simple Demographics Model, which was formulated from the most current WAIS-IV standardization sample, the corrected Barona equation performed no differently than the TOPF Simple Demographics model. This effect was observed across differing racial and age groups. This finding indicates that, in terms of demographic models, the corrected Barona equation is a viable option for predicting premorbid IQ and may be a more economical choice for clinicians since the 1984 original equation is freely available (Barona et al., 1984) as is the meta-analysis by which the .23 IQ point correction was calculated (Trahan et al., 2014). Furthermore, although there was not significant difference between the TOPF simple demographics and the adjusted Barona equation, the adjusted Barona had a lower mean difference than the TOPF to actual FSIQ and predicted a higher percentage of the variance in IQ.

The primary limitation of this study was that by using a clinical sample from an outpatient neuropsychology clinic, only 8% did not meet criteria for any *DSM–5* diagnoses and over half of the cognitively unimpaired sample met criteria for a psychiatric diagnosis. One way this limitation was addressed was to include only those participants who passed measures of performance validity; however, this does not guarantee that our measurement of current intellectual functioning was not at all impacted by mood. Another limitation is the restricted range of WAIS-IV FSIQ scores in the sample (i.e., 77–125); however, although not as wide a range as the standardization sample, this sample does represent a broad ability range. Because of this, we were not able to examine the accuracy of the prediction for individuals whose intellectual abilities were extremely below or above the mean (i.e., ≥2 *SD*s); however, traditionally demographic models have been rather poor at predicting the extremes of the distribution (Helmes, 1996; Veiel & Koopman, 2001). Because of the availability of WAIS-IV scores without accompanying WAIS-III scores, the most accurate OPIE-3 scores could not be calculated, and an appropriate adjustment for the Flynn effect could not be applied and tested. Although the relative size of the male sample to the female portion of the sample contributes to a sex-based biasing of the results, prior research shows that sex (although a significant contributor of variance) explains relatively less variance (refer to the relative weights of the variable in Footnote 1) in IQ than other demographic factors, particularly race, education, and occupations (Barona et al., 1984). Finally, although the sample was diverse, both racially and linguistically, the sample size for some of these subgroups was relatively small. This may limit the generalizability of some findings to other populations; however, even in the context of these smaller subgroups, findings were consistent with prior studies.

In sum, demographic models can provide accurate estimates of premorbid IQ when the Flynn effect is adequately addressed (e.g., corrected Barona equation) or when they are derived from the most updated WAIS standardization sample (e.g., TOPF Simple Demographics Model). These are important tools for clinicians, given that performance-based or word reading models may not always be the most prudent option when estimating premorbid IQ. Demographic models are not confounded by neurological, developmental, or physical impairment that would affect reading ability or cognitive performance, nor are they subject to psychometric issues of multicollinearity which would impact performance-based measures. Further, demographic models offer an efficient, quick method of estimating premorbid intelligence without the need for extensive additional testing. This study demonstrated that the Flynn-corrected Barona equation is a free, viable option for estimating premorbid intellectual functioning and provides clinicians with options, depending on the individual demands and goals of the assessments. Future studies may examine further these models in more diverse or older

adult samples, particularly those where irregular word reading is likely to be less feasible.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Ashendorf, L., Jefferson, A. L., Green, R. C., & Stern, R. A. (2009). Test-retest stability on the WRAT-3 reading subtest in geriatric cognitive evaluations. *Journal of Clinical and Experimental Neuropsychology, 31,* 605–610. http://dx.doi.org/10.1080/13803390802375557

Axelrod, B. N., Vanderploeg, R. D., & Schinka, J. A. (1999). Comparing methods for estimating preformatted intellectual functioning. *Archives of Clinical Neuropsychology, 14,* 341–346. http://dx.doi.org/10.1016/S0887-6177(98)00028-6

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS–R. *Journal of Consulting and Clinical Psychology, 52,* 885–887. http://dx.doi.org/10.1037/0022-006X.52.5.885

Batchelor, E., & Meyers, J. (2013). A-81 Estimating WAIS-IV FSIQ using the Barona and Ward 7 Short Form. *Archives of Clinical Neuropsychology, 28,* 518–626.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B.* Methodological, *57,* 289–300. http://dx.doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23,* 729–741. http://dx.doi.org/10.1080/13854040802427803

Carlozzi, N. E., Stout, J. C., Mills, J. A., Duff, K., Beglinger, L. J., Aylward, E. H., . . . the PREDICT-HD Investigators of the Huntington Study Group. (2011). Estimating premorbid IQ in the prodromal phase of a neurodegenerative disease. *The Clinical Neuropsychologist, 25,* 757–777. http://dx.doi.org/10.1080/13854046.2011.577811

Cockburn, J., Keene, J., Hope, T., & Smith, P. (2000). Progressive decline in NART score with increasing dementia severity. *Journal of Clinical and Experimental Neuropsychology, 22,* 508–517. http://dx.doi.org/10.1076/1380-3395(200008)22:4;1-0;FT508

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: An emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist, 31,* 1029–1046. http://dx.doi.org/10.1080/13854046.2017.1323119

Farcomeni, A. (2006). More powerful control of the false discovery rate under dependence. *Statistical Methods and Applications, 15,* 43–73. http://dx.doi.org/10.1007/s10260-006-0002-z

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191. http://dx.doi.org/10.1037/0033-2909.101.2.171

Fogel, M. L. (1964). The intelligence quotient as an index of brain damage. *American Journal of Orthopsychiatry, 34,* 555–562. http://dx.doi.org/10.1111/j.1939-0025.1964.tb02225.x

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology, 12,* 711–738. http://dx.doi.org/10.1093/arclin/12.8.711

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology, 67,* 850–857. http://dx.doi.org/10.1016/j.jclinepi.2014.03.012

Helmes, E. (1996). Use of the Barona method to predict premorbid intelligence in the elderly. *Clinical Neuropsychologist, 10,* 255–261. http://dx.doi.org/10.1080/13854049608406688

Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *Clinical Neuropsychologist, 9,* 83–88. http://dx.doi.org/10.1080/13854049508402063

Ladd, C. E. (1964). WAIS performances of brain damaged and neurotic patients. *Journal of Clinical Psychology, 20,* 114–117. http://dx.doi.org/10.1002/1097-4679(196401)20:1<114::AID-JCLP2270200116>3.0.CO;2-J

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *Clinical Neuropsychologist, 17,* 410–425. http://dx.doi.org/10.1076/clin.17.3.410.18089

Larrabee, G. J., Largen, J. W., & Levin, H. S. (1985). Sensitivity of age-decline resistant ("hold") WAIS subtests to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 7,* 497–504. http://dx.doi.org/10.1080/01688638508401281

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York, NY: Oxford University Press.

Maddrey, A. M., Cullum, C. M., Weiner, M. F., & Filley, C. M. (1996). Premorbid intelligence estimation and level of dementia in Alzheimer's disease. *Journal of the International Neuropsychological Society, 2,* 551–555. http://dx.doi.org/10.1017/S1355617700001727

McFie, J. (1975). *Assessment of organic intellectual impairment*. Oxford, UK: Academic Press.

Mortensen, E. L., Gade, A., & Reinisch, J. M. (1991). A critical note on Lezak's 'best performance method' in clinical neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 13,* 361–371. http://dx.doi.org/10.1080/01688639108401050

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART)*. Windsor, UK: Nfer-Nelson.

Norton, K., Watt, S., Gow, B., & Crowe, S. F. (2016). Are tests of premorbid functioning subject to the Flynn effect? *Australian Psychologist, 51,* 374–379. http://dx.doi.org/10.1111/ap.12235

O'Rourke, J. J., Adams, W. H., Duff, K., Byars, J., Nopoulos, P., Paulsen, J. S., & Beglinger, L. J. (2011). Estimating premorbid functioning in Huntington's disease: The relationship between disease progression and the wide range achievement test reading subtest. *Archives of Clinical Neuropsychology, 26,* 59–66. http://dx.doi.org/10.1093/arclin/acq088

Pearson. (2009). *Advanced Clinical Solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual*. San Antonio, TX: Pearson.

Reynolds, C. R., & Gutkin, T. B. (1979). Predicting the premorbid intellectual status of children using demographic data. *Clinical Neuropsychology, 1,* 36–38.

Schoenberg, M. R., Scott, J. G., Duff, K., & Adams, R. L. (2002). Estimation of WAIS-III intelligence from combined performance and demographic variables: Development of the OPIE-3. *The Clinical Neuropsychologist, 16,* 426–438. http://dx.doi.org/10.1076/clin.16.4.426.13913

Schoenberg, M. R., Duff, K., Scott, J. G., & Adams, R. L. (2003). An evaluation of the clinical utility of the OPIE-3 as an estimate of premorbid WAIS-III FSIQ. *The Clinical Neuropsychologist, 17,* 308–321. http://dx.doi.org/10.1076/clin.17.3.308.18088

Schoenberg, M. R., Duff, K., Scott, J. G., Patton, D., & Adams, R. L. (2006). Prediction errors of the Oklahoma Premorbid Intelligence Estimate-3 (OPIE-3) stratified by 13 age groups. *Archives of Clinical Neuropsychology, 21,* 469–475. http://dx.doi.org/10.1016/j.acn.2006.06.006

Schroeder, R. W., Martin, P. K., Heinrichs, R. J., & Baade, L. E. (2018). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist, 33,* 466–477.

Smith-Seemiller, L., Franzen, M. D., Burgess, E. J., & Prieto, L. R. (1997). Neuropsychologists' practice patterns in assessing premorbid intelligence. *Archives of Clinical Neuropsychology, 12,* 739–744. http://dx.doi.org/10.1093/arclin/12.8.739

Spinks, R., McKirgan, L. W., Arndt, S., Caspers, K., Yucuis, R., & Pfalzgraf, C. J. (2009). IQ estimate smackdown: Comparing IQ proxy measures to the

WAIS-III. *Journal of the International Neuropsychological Society, 15,* 590–596. http://dx.doi.org/10.1017/S1355617709090766

Taylor, R. (2000). National Adult Reading Test performance in established dementia. *Archives of Gerontology and Geriatrics, 29,* 291–296. http://dx.doi.org/10.1016/S0167-4943(99)00042-4

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140,* 1332–1360. http://dx.doi.org/10.1037/a0037173

Vanderploeg, R. D., Schinka, J. A., & Axelrod, B. N. (1996). Estimation of WAIS–R premorbid intelligence: Current ability and demographic data used in a best-performance fashion. *Psychological Assessment, 8,* 404–411. http://dx.doi.org/10.1037/1040-3590.8.4.404

Vanderploeg, R. D., & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid abiltiy: Combining subtest performance and demographic variable predictors. *Achives of Clinical Neuropsychology, 10,* 225–239. http://dx.doi.org/10.1093/arclin/10.3.225

Veiel, H. O., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment, 13,* 356–368. http://dx.doi.org/10.1037/1040-3590.13.3.356

Webber, T. A., Critchfield, E. A., & Soble, J. R. (in press). Convergent, discriminant, and concurrent validity of non-memory-based performance validity tests. *Assessment*.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale (WAIS)*. New York, NY: Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised (WAIS–R)*. New York: Psychological Corporation.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale,* 3rd ed. (*WAIS-III*). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2001). *Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *WAIS-IV: Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 (WRAT4)*. Lutz, FL: Psychological Assessment Resources, Inc.

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: A cross-validation. *Journal of Clinical and Experimental Neuropsychology, 13,* 204–216. http://dx.doi.org/10.1080/01688639108401038

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology, 46,* 1554–1555. http://dx.doi.org/10.1037/0022-006X.46.6.1554

Zhang, J., & Coombes, K. R. (2012). Sources of variation in false discovery rate estimation include sample size, correlation, and inherent differences between groups. *BMC Bioinformatics, 13*(Suppl. 13), S1. http://dx.doi.org/10.1186/1471-2105-13-S13-S1

Received December 21, 2018
Revision received April 2, 2019
Accepted April 27, 2019 ■